

Análisis Filogenético

Algoritmos en Bioinformática

Luis E. Garreta U

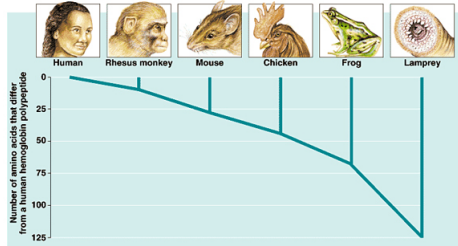
luis.garreta@javerianacali.edu.co

Doctorado en Ingeniería
Pontificia Universidad Javeriana – Cali
Facultad de Ingeniería - Carrera de Biología

11 de abril de 2018

Introducción

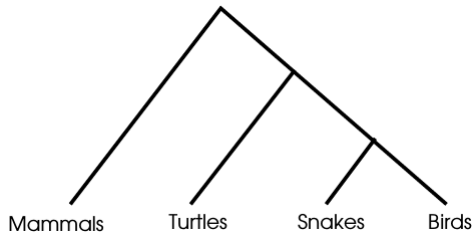
- ▶ Los árboles filogenéticos muestran las relaciones evolutivas entre grupos de organismos o familias de secuencias de proteínas o nucleóticos
- ▶ Pregunta: Cómo una familia pudo haberse derivado durante la evolución?



Copyright © Pearson Education, Inc., publishing as Benjamin Cummings.

Árbol Hipotético que relaciona los organismos

El principal objetivo de construir árboles filogenéticos es ilustrar las relaciones evolutivas entre grupos de organismos.



Aplicaciones del Análisis Filogenético

- ▶ Analizar cambios ocurridos en la evolución de los diferentes organismos
- ▶ Estimar el tiempo de divergencia entre organismos a partir de su ultimo ancestro compartido
- ▶ Predecir que genes tienen funciones similares.
- ▶ Seguir cambios que ocurren rápidamente en especies (ej. virus HIV).

Tipos de Datos usados para la construcción de los árboles

► Tradicionalmente: características morfológicas:

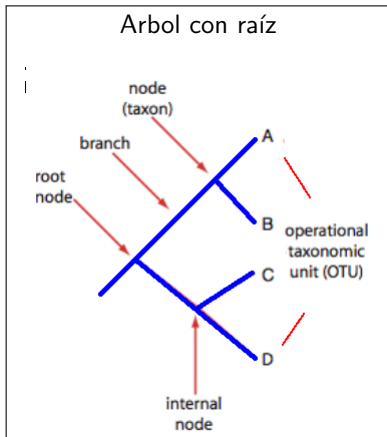
- ✓ Presencia de plumas,
- ✓ Numero de patas,
- ✓ Forma del pico,
- ✓ ...

► **Actualmente:** Datos Moleculares:

- ✓ Secuencias de proteínas
- ✓ Secuencias de ADN
- ✓ Secuencias de ARN
- ✓ Otros tipos de secuencias:

- Tablas de características (ej. #Patas, Tipo Pico, Cola)

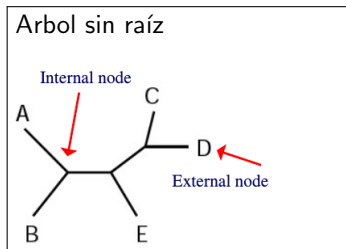
Definiciones y Terminología Árboles Filogenéticos



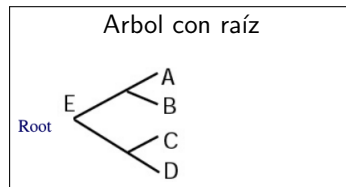
- ▶ **Nodos externos (Hojas):** Elementos bajo comparación: *OTUs*^a
- ▶ **Nodos internos:** Unidades ancestrales hipotéticas. Agrupan a las OTUs
- ▶ **Raíz:** Ancestro común de todas las OTUs
- ▶ **Camino evolutivos:** Camino desde la raíz a una OTU específica
- ▶ **Topología:** Forma (posición) de las ramas dentro del árbol
- ▶ **Longitud de rama:** Número de cambios, distancia genética, tiempo evolutivo.

^a*Operational Taxonomic Units*

Tipos de Árboles



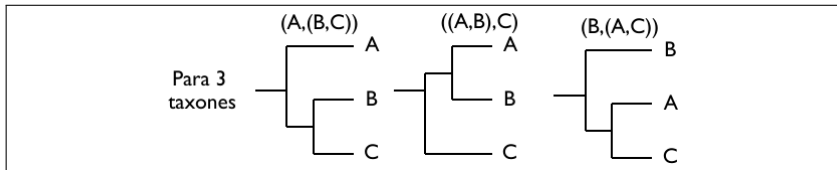
→



Podemos adicionar un elemento muy alejado evolutivamente (**outgroup**) a un árbol sin raíz

Y obtenemos un árbol con raíz donde **E** es el ancestro más común de todos los elementos (OTUs)

Número de Póssibles Árboles



Árboles sin raíz:

- ▶ Para n OTUs, existen $\prod_{i=3}^n (2i - 5)$ árboles sin raíz.
 - ✓ 3 OTUs : 1 árbol
 - ✓ 5 OTUs : 15 árboles
 - ✓ 10 OTUs : 2,027,025

Árboles con raíz:

- ▶ Para n OTUs, existen $\prod_{i=3}^n (2i - 3)$ árboles con raíz.
 - ✓ 3 OTUs : 3 árboles
 - ✓ 5 OTUs : 105 árboles
 - ✓ 10 OTUs : 34,459,425

Construcción de Arboles Filogenéticos

► Métodos de distancia:

- ✓ Construye el árbol basado en la distancia evolutiva para todas las OTUs

► Máxima Parsimonia:

- ✓ Construye un árbol que minimice el número de cambios requeridos para explicar los datos

► Máxima Similitud:

- ✓ Construye un árbol que maximiza la probabilidad de ser el generador de las secuencias observadas.

Construcción de Árboles por Métodos de Distancia

Pasos:

1. Calcular Matriz de Distancias

- 1.1 Alinear las secuencias

- 1.2 Calcular las mutaciones

- 1.3 Crear una matriz de distancias

2. Agrupar los elementos usando por ejemplo UPGMA^a

- 2.1 Buscar en la matriz el par de elementos más cercano

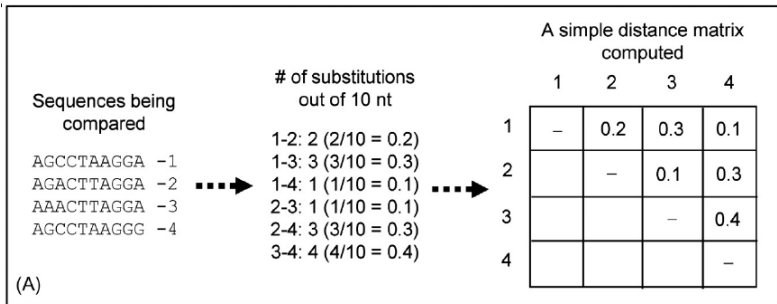
- 2.2 Agrupar ese par y recalcular distancias

- 2.3 Volver a ejecutar los pasos desde 4 hasta agrupar a todos los elementos

^aAlgoritmo de Agrupamiento Secuencial sin Pesos

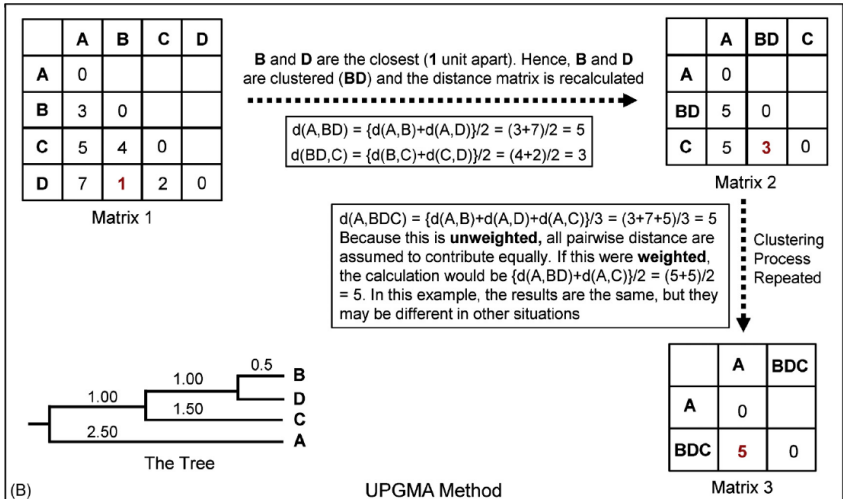
Métodos de Distancia:

Paso1: Calcular la Matriz de Distancias



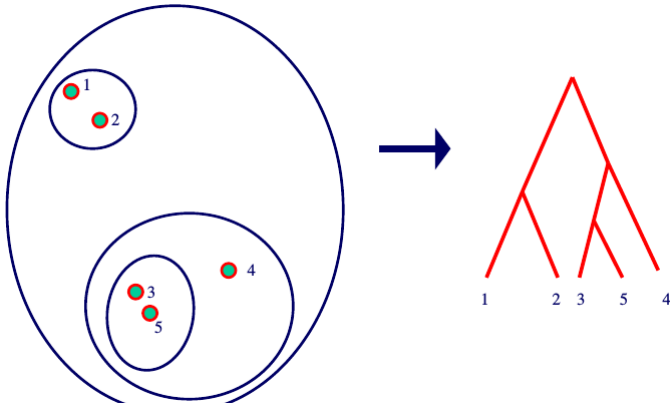
Métodos de Distancia:

Paso2: Agrupar los elementos usando por ejemplo UPGMA



Algoritmo de Agrupamiento UPGMA

- ▶ UPGMA es un algoritmo de agrupamiento secuencial jerárquico
- ▶ Se basa en agrupar los pares con distancias más cercanos
- ▶ Cada par se convierte en un nuevo grupo y este grupo se convierte en un nuevo elemento



Ejemplo2: Construir el Árbol Filogenético

- Suponga que se tienen cuatro secuencias y su matriz de distancias es la siguiente:

	A	B	C	D
A	0			
B	8	0		
C	7	9	0	
D	12	14	11	0

e

Debilidades UPGMA

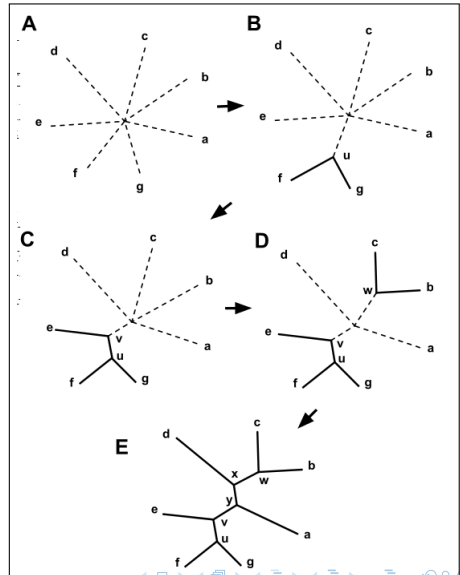
- ▶ UPGMA asume que tasas de evolución son iguales para todos los nodos
 - ✓ El reloj molecular de todos los nodos es igual
- ▶ A menos que lo anterior sea verdad para su problema, entonces use UPGMA
- ▶ Sino, mejor usar NEIGHBOR JOINING

Método de Agrupamiento de Neighbor Joining

- ▶ Método basado en distancias más usado para reconstrucción de árboles filogenéticos
- ▶ Idea General:
 - ✓ También tenga en cuenta las distancias a los otros nodos

Idea General Neighbor Joining

1. Comenzamos con todos los OTUs unidos directamente en un árbol sin raíz (todos son vecinos)
2. Se hacen las $N(N-1)/2$ comparaciones entre OTUs vecinos para determinar cuál es la pareja más cercana
3. Esos OTUs se unen mediante un nuevo nodo interno y volvemos al paso dos, decrementando en 1 el valor de N



Limitaciones Neighbor Joining

- El algoritmo minimiza la longitud de una rama en cada paso, así que no asegura una longitud mínima global (algoritmo voráz)

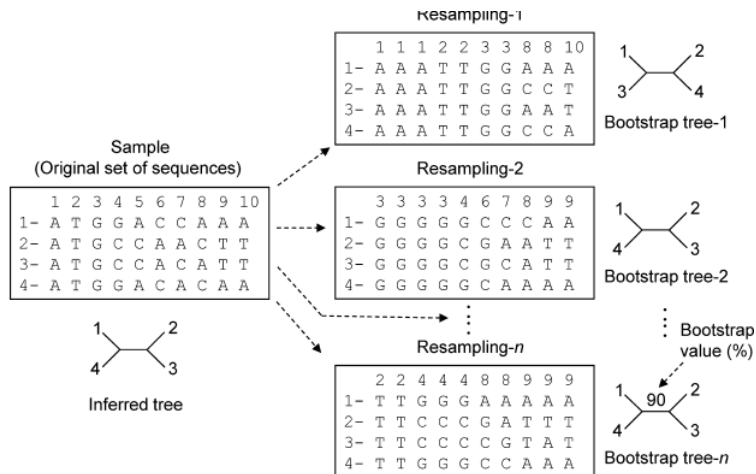
Evaluación de los árboles

- ▶ Que un programa informático produzca un árbol filogenético no significa que sea correcto:
 - ✓ GIGO (Garbage In, Garbage Out)
- ▶ En muchos casos puede ser globalmente correcto pero tener inexactitudes en algunas ramas
- ▶ Evaluación: bootstrapping o remuestreo:
 - ✓ Verificación del significado biológico de un árbol evaluando su robustez

Bootstrapping

- ▶ Primero, seleccionamos columnas del AMS original de forma aleatoria, hasta tener tantas como en el AMS original
 - ✓ Se permiten repeticiones (muestreo con reemplazamiento)
 - ✓ Es un alineamiento artificial, pero que conserva las características del AMS original
 - ✓ Se realizan muchos de estos muestreos aleatorios (100 a 1000)
- ▶ A cada AMS aleatorio se le aplica el algoritmo a evaluar, obteniendo un árbol
- ▶ Se construye un árbol de consenso con todos los árboles obtenidos:
 - ✓ El porcentaje de veces que una ramificación aparece es el valor de bootstrap
 - ✓ Valores de bootstrap $> 70\%$ suelen tomarse como suficientemente robustos (equivalen a un nivel de significatividad $p < 0.05$)

Idea Bootstrapping



Herramientas para Análisis Filogenético

- ▶ Tools:

- ✓ MEGA

- ▶ Online Servers:

- ✓ Reconstructing and analysing phylogenetic relationships between molecular sequences.

- ▶ <http://www.phylogeny.fr/>

- ✓ Simple Phylogeny (EBI):

- ▶ https://www.ebi.ac.uk/Tools/phylogeny/simple_phylogeny/